

# Boolean Compressed Sensing and Noisy Group Testing

George Atia and Venkatesh Saligrama

## Abstract

The fundamental task of group testing is to recover a small distinguished subset of items from a large group while efficiently reducing the total number of tests (measurements). The key contribution of this paper is in adopting a new information-theoretic perspective on group testing problems. Establishing its connection to Shannon-coding theory, we formulate the group testing problem as a channel coding/decoding problem and derive a unifying result that reduces many of the interesting questions to computation of a mutual information expression. First, we derive an achievable bound on the total number of tests using random constructions and typical set decoding. This result is fairly general; it allows us to verify existing bounds for some of the known scenarios and extend the analysis to many new interesting setups including noisy versions of group testing. We obtain tradeoffs between the number of tests, sparsity and the group size for a wide range of models. Among the models we consider are the deterministic noise-free case, approximate reconstruction with bounded distortion, additive measurement noise, and dilution models. Then, we show that the achievable bound provides fairly tight asymptotic scaling by deriving an information theoretic lower-bound using Fano's inequality.

## I. INTRODUCTION

Group testing has been effectively used in numerous applications. It was originally proposed during World War II to reduce the total number of blood tests required to detect soldiers with Syphilis [2], [3]. Instead of conducting a separate blood test for each and every soldier, the idea was to pool blood samples from many soldiers and test them simultaneously. Group testing has been used in general in biology for screening libraries of DNA clones (strings of DNA sequence) of the human genome and for screening blood for diseases. Other applications include quality control for detecting defective parts in production

This work was presented in part at the Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2009 [1]. Prior to that, a preprint was made publicly available on arXiv in July 2009. This work was done at the Department of Electrical and Computer Engineering, Boston University, Boston, MA, Emails: {geokamal,srv}@bu.edu.

lines, data forensics to test collections of documents by applying one-way hash functions, computer fault diagnosis, and contention algorithms in multiple access communications. Recently, group testing methods have also been applied to spectrum enforcement in cognitive radios [4], [5].

In the basic group testing problem [2], we are given a population of  $N$  items. Among them, at most  $K$  items, also called defectives, are of interest. The set of defectives is denoted by  $\mathcal{G}$ . Associated with the group testing problem is a binary matrix  $M$  known as the measurement matrix. This matrix defines the assignment of each of the items to different pools or collections. The  $(i, j)$  entry is 1 if the  $j$ -th item is contained in the  $i$ -th pool and 0 otherwise. A test conducted on the pool is positive if there is at least one item belonging to the pool which is also an element of  $\mathcal{G}$ , and is negative otherwise. A measurement matrix for a non-adaptive group testing algorithm is a  $T \times N$  matrix where the rows correspond to pools of items and  $N$  the number of items. Each item is manifested as a vector codeword of length  $T$ . If  $K$  items are defective, then the  $T$  tests are a boolean sum of  $K$  columns of the measurement matrix. The goal is to construct a pooling design to recover the defective set while reducing the required number of tests. Group testing is related in spirit to compressed sensing (CS). In CS we are given an  $N$ -dimensional sparse signal with support size  $K$ . Random projections of the sparse signal are obtained. The goal is to identify the support set while minimizing the number of projections.

While the degradation of CS with noise has been characterized (see [6] and references therein), the noisy group testing problem, the main focus of this paper, has not been systematically studied. Our attention in this paper is on the so called non-adaptive testing problem [2], where the measurement matrix is formed prior to performing the tests.

A significant part of the existing research on group testing is focused on combinatorial pool design (i.e. construction of measurement matrices) to guarantee the detection of the items of interest using a small number of tests. Two types of matrix constructions have been considered. Disjunct matrices [2] satisfy the so called covering property<sup>1</sup>. In the context of group testing this property implies that a test pattern obtained by taking any  $K$  columns of the measurement matrix does not cover any other boolean sum of  $K$  or smaller number of columns. Matrices that satisfy this property are often referred to as superimposed codes and combinatorial constructions were extensively developed by [7], [8], [9]. Superimposed codes are not only desirable because they ensure identifiability but they also lead to efficient decoding. Separability [2] is a weaker notion that is also often employed. A separable matrix ensures that the boolean sum of  $K$  columns are all distinct, which ensures identifiability. Uniquely decipherable

<sup>1</sup>We say that a column  $x$  is covered by a column  $y$  iff  $x \vee y = y$ .

codes [8], [7] are codes that guarantee that every boolean sum of  $K$  or smaller number of columns are distinct. Recently, it has been shown [10] that all of these notions are equivalent upto a scaling factor on the number of tests  $T$ .

A different approach to group testing based on probabilistic method has also been advocated by several researchers [11], [12], [13], [14], [15]. Dyachkov and Rykov [11] and Ruszinko [12] developed upper and lower bounds on the number of rows  $T$  for a matrix to be  $K$ -disjunct (bound on length of superimposed codes). Random designs were used to compute upper bounds on the lengths of superimposed codes by investigating when randomly generated matrices have the desired covering/separability properties. They showed that for  $N \rightarrow \infty$  and  $K \rightarrow \infty$ ,  $T \geq O(\frac{K^2 \log N}{\log K})$  for exact reconstruction with worst-case input. Sebo [13] investigated average error probabilities and showed that for an arbitrarily small error probability, a randomly generated matrix will be  $K$ -disjunct if  $T = O(K \log N)$  as  $N \rightarrow \infty$ . Recently Berger et al. [14] proved upper and lower bounds on the number of tests for two-stage disjunctive testing. Approximate reconstruction [15], whereby a fraction  $\alpha$  of the defective items are allowed to be in error, has been described. Again the number of tests here has been shown to scale as  $T = O(K \log N)$  as  $N \rightarrow \infty$ .

While these approaches have generally characterized fundamental tradeoffs for noiseless group testing, the noisy counterpart of group testing has not been systematically addressed. In this paper we take a novel information theoretic approach to group testing problems. The common approach to previous related work was to prove bounds on the size of randomly generated matrices to exhibit the aforementioned separability and covering properties. In contrast, we formulate the problem as a detection problem and establish its connection to Shannon coding theory [16] which, to the best of our knowledge, is explored in this paper for the first time. While there exists a one-to-one mapping between both formulations, the new perspective allows us to easily obtain results for a wide range of models including noisy versions of group testing. Our approach, which is fairly general, is to map the group testing problem to a corresponding channel model which allows the computation of simple mutual information expressions to derive achievable bounds on the required number of tests. This result allows us to verify existing bounds for some of the known scenarios and extend the analysis to many new interesting setups including noisy versions of group testing.

One major contribution of this paper is to develop tools to analyze long standing noisy versions of the group testing problem. In particular we consider two models: The dilution model and the additive model.

- Additive model: False alarms could arise from errors in some of the screening tests. This is manifested when some tests are erroneously positive.

Model	$T$ (average $P_e$ criterion)	$T$ (max $P_e$ criterion)
Noise-free	$O(K \log N)$	$O(K^2 \log N)$
Reconstruction with distortion	$O(K \log N)$	$O(K \log N)$
Noisy with additive noise	$O(\frac{K \log N}{1-q})$	$O(\frac{K^2 \log N}{1-q})$
Noisy with dilution	$O(\frac{K \log N}{(1-u)^2})$	$O(\frac{K^2 \log N}{(1-u)^2})$

TABLE I

THIS TABLE SUMMARIZES THE SCALING RESULTS FOR THE VARIOUS MODELS CONSIDERED IN THE PAPER. IN PARTICULAR, IT SHOWS THE REQUIRED NUMBER OF TESTS  $T$  AS A FUNCTION OF THE SIZE OF THE DEFECTIVE SET  $K$ , THE TOTAL NUMBER OF ITEMS  $N$ , AND THE MODEL PARAMETERS FOR THE NOISELESS, ADDITIVE NOISE, AND DILUTION MODELS FOR BOTH THE AVERAGE ERROR AND WORST-CASE ERROR CRITERIA.

- Dilution model: Even though a positive item is contained in a given pool, the test's outcome could be negative if the item's presence gets diluted for that specific test. For example, in blood testing the positive sample might get diluted in one or more tests leading to potential misses of infected blood samples. To account for such a case we analyze the group testing problem where some of the positive entries get flipped into zeros with a given probability  $u$ .

Table I summarizes the scaling results we obtained for the various models considered in the paper for the average error and worst case error criteria.

The rest of the paper is organized as follows. Section II describes the problem setup. The main achievable result mapping the problem to mutual information expression is provided in Section III. Section IV considers the noise free (deterministic) version of the problem with average and worst case errors. Approximate reconstruction with distortion is investigated in Section V. In Section VI we consider different noisy models with additive and dilution effects. In Section VII we prove a converse bound using Fano's inequality. Finally, we provide conclusions in Section VIII.

## II. PROBLEM SETUP

Among a population of  $N$  items,  $K$  unknown items are of interest. The collection of these  $K$  items represents the defective set. The goal is to construct a pooling design, i.e. a collection of tests, to recover the defective set while reducing the number of required tests. The idea is illustrated in Fig. 1. In this example, the defective set  $\mathcal{G} = \{2\}$ , i.e.,  $K = 1$ , since only the second item is defective. The shown binary valued matrix, represents the measurement matrix (transposed) defining the assignment of items to

tests. The entry is 1 if the item is a member of the designated test and 0 otherwise. At the bottom of the figure we highlight the positive tests. The outcome of a test is positive if and only if the second item is a member of that test. Observing the output for a number of tests  $T$ , the goal is to recover the defective set  $\mathcal{G}$ . While in combinatorial group testing the goal is to find the defective set for the worst-case input, probabilistic group testing requires the average error to be small. Both formulations are considered in this paper in subsequent sections. We assume that the item-test assignment is generated randomly. Before we provide our main result we define the following notation:

- $N$ : Total number of items
- $K$ : Known number of defectives (or positive items)
- $p$ : Probability that an item is part of a given test
- $T$ : Total number of tests
- $X_j^T$ : for the  $j$ -th item,  $X_j$  is a binary vector  $\in \{0, 1\}^T$ , with the  $t$ -th entry  $X_j(t) = 1$  if the  $j$ -th item is pooled in test  $t$ , and 0 otherwise. Following information theoretic convention, we call it the  $j$ -th codeword since we will be using an information theoretic framework. Upper-case letters are used to denote random variables and vectors, and lower-case letters for denoting their realizations, e.g.  $X_j$  is a random vector and  $x_j$  a realization of this random vector.
- $Y^T$  is the binary observation vector of length  $T$ , with entries = 1 for the tests with positive outcome. Similarly,  $y$  denotes a realization of  $Y$ .
- $S_\ell \subset \{1, 2, \dots, N\}$  denotes a subset of items of size  $\ell$ .
- $X_{S_\ell}^T$  denotes the collection of codewords of length  $T$  corresponding to the items in  $S_\ell$ . It is a matrix of size  $\ell \times T$ . We drop the  $T$  when we refer to an instance, i.e., a vector of code symbols. To avoid cumbersome notation, for the rest of the paper we use the shorthand notation  $X_{(\ell)}$  instead of  $X_{S_\ell}$ . It should be of no confusion that we are referring to the codewords of a specific subset of items.
- $\mathcal{C}$ : The measurement matrix, or the codebook, i.e., a collection of  $N$  codewords defining the pool design, that is, the assignment of items to tests.

#### Noise-free case:

For the noise-free case, the outcome of the tests  $Y$  is deterministic. It is the boolean sum of the codewords corresponding to the defective set  $\mathcal{G}$ . In other words:

$$Y = \bigvee_{i \in \mathcal{G}} X_i. \quad (1)$$

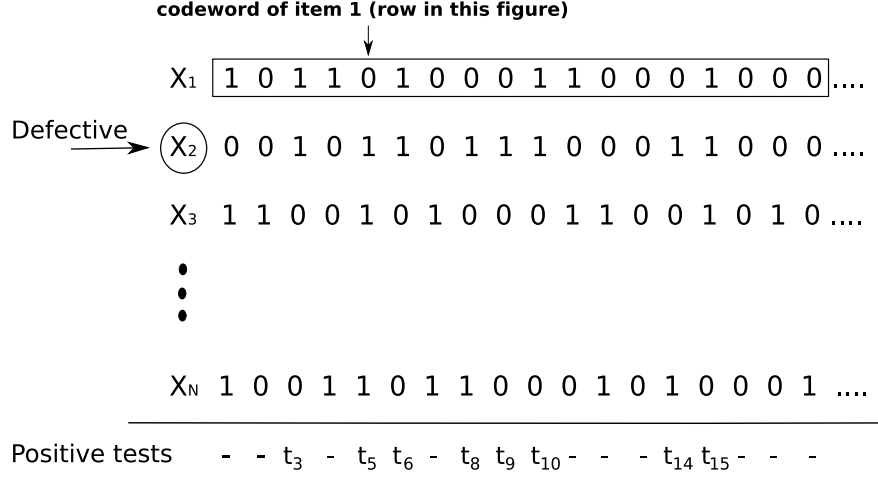


Fig. 1. The shown binary structure defines the assignment of items to tests. Rows are codewords for corresponding items, and columns correspond to pools of items (tests). The entry is 1 if the item is a member of the designated test and 0 otherwise. In the shown example, the defective set  $\mathcal{G} = \{2\}$ , i.e. only the second item is defective. At the bottom of the figure we show the positive tests. The outcome of a test is positive if and only if the second item is a member of that test. Observing the output for a number of tests  $T$ , the hope is to recover the defective set  $\mathcal{G}$ .

Alternatively, if  $R_i \in \{0, 1\}$  is an indicator function for the  $i$ -th item determining whether it belongs to the defective set, i.e.,  $R_i = 1$  if  $i \in \mathcal{G}$  and  $R_i = 0$  otherwise, then the outcome  $Y_t$  of the  $t$ -th test in the noise-free case can be written as:

$$Y_t = \bigvee_{i=1}^N X_{ti} R_i \quad (2)$$

where  $X_{ti}$  is the binary entry at cell  $(t, i)$  of the measurement matrix  $X$ .

#### Noisy cases:

In this paper we also consider two noisy models, the additive model and the dilution model.

- **Additive Model:** In this model we account for false alarms in the outcome of pooling tests. The outcome of a test can still be 1 even if no positive items are pooled in that test. This effect is captured by adding independent bernoulli( $q$ ) random variables  $W_t$  to the outcome of the  $t$ -th test of the noise-free model in Eq.(2), i.e.,:

$$Y_t = \left( \bigvee_{i=1}^N X_{ti} R_i \right) \vee W_t \quad (3)$$

where  $W_t \sim B(q), t = 1 \dots T$ .

- **Dilution Model:** Positive items taking part in a given test might probabilistically behave as absent (diluted). If all positive items in a given test appear as absent, that could potentially lead to erroneously zero outcomes. This model is motivated by blood dilution due to pooling with other negative tests or imperfectly diluted blood samples. This effect is captured by the Z-channel model of Fig. 2. The outcome of the  $t$ -th test can be written as:

$$Y_t = \bigvee_{i=1}^N \mathcal{Z}(X_{ti}R_i) \quad (4)$$

where  $\mathcal{Z}$  represents the Z-channel model of Fig. 2.

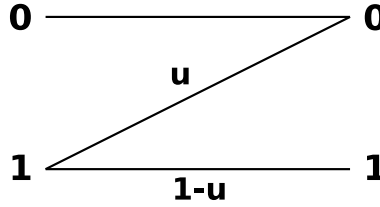


Fig. 2. Dilution channel (Z-channel): positive items taking part in a given test might probabilistically behave as absent (diluted). In other words, even though a positive item is contained in a given pool, the test's outcome could be negative if the item's presence gets diluted for that specific test. For example, in blood testing the positive sample might get diluted in one or more tests leading to potential misses of infected blood samples.

Note that in the additive model, the outcome of testing a pool with no defective items might be erroneously positive, i.e., false positives would occur. On the other hand, the membership of a defective item in a given test might go unnoticed in the dilution model. If all defectives appearing in a given pool are diluted, a false negative occurs. The effect of dilution on the increase in number of tests is expected to be more severe than the effect of additive noise. This is explained by the fact that tests with negative outcomes are generally more informative; while a truly positive test merely indicates that at least one defective is present in the pool, a truly negative test exonerates all the members pooled in that test. With additive noise, a test with a negative outcome is never erroneous. In contrast, dilution diminishes our confidence in pools with negative outcomes, since a seemingly negative outcome would not necessarily mean all pooled members are perfect. Intuitively speaking, additive noise can be potentially mitigated by repetition of tests whereas the dilution effect is more intricate to resolve. This intuition is verified by the results we obtained through theoretical analysis as shown in the next sections.

For notation, random vectors and their realizations are denoted by upper- and lower-case letters, respectively. Index collections of codewords of size  $K$  by  $V$ , where  $V \in \{1, 2, \dots, \binom{N}{K}\}$ . Defining a decoding function  $g(\cdot) : Y^T \rightarrow V$ ,  $g(\cdot)$  maps the outcome of the tests  $Y$  of length  $T$  to an index  $V$  corresponding to a specific set of defectives. Now define the error  $\lambda_v$  as:

$$\lambda_v = \sum_{y^T} p(y^T | x_v^T) \mathcal{I}(g(y^T) \neq v | v) \quad (5)$$

where  $\mathcal{I}$  is an indicator function.  $\lambda_v$  is the probability of error conditioned on a given  $v$ , i.e. the probability that the decoded set is different from the true defective set. Note that for the deterministic noise-free case this simplifies to  $\lambda_v = \mathcal{I}(g(y^T) \neq v | v)$ . Averaging over all possible inputs  $v$ , we define the average error  $\lambda$  as:

$$\lambda = \frac{1}{\binom{N}{K}} \sum_v \lambda_v \quad (6)$$

For the aforementioned models (noiseless and noisy), we prove achievable and converse bounds on the total number of tests  $T$  as  $N, K \rightarrow \infty$ . Namely, we consider the following criteria:

- Criteria 1: Average error probability:

$$\lambda_{avg} = \frac{1}{\binom{N}{K}} \sum_v \lambda_v$$

- Criteria 2: Worst-case error probability:

$$\lambda_{\max} = \max_v \lambda_v$$

- Criteria 3: Reconstruction with distortion:

In this case, we are satisfied with approximate reconstruction of the defective set. Let  $d$  be a distance function between the decoded set  $g(y^T)$  and the true defective set  $v$  such that  $d(g(y^T), v)$  is equal to the number of misses. Given  $K$  declared candidates, an error occurs only if the number of missed items is greater than  $\alpha K$ , i.e.,

$$\lambda_v = \mathcal{I}(d(g(y^T), v) > \alpha K | v)$$

In the following section we derive our main result. We prove an achievable bound on the number of tests  $T$  where  $T(N, K)$  is said to be achievable if  $\forall \delta > 0$  there exists a pool design of size  $T \times N$  such that  $\lambda < \delta$ , where  $\lambda$  is defined according to the performance criteria defined above. The bound is fairly general as it applies to the noise-free and the noisy versions of the problem as we elaborate in the following sections.



### III. MAIN RESULT: ACHIEVABLE BOUND

To derive an achievable bound on the number of tests  $T$ , we show how the group testing problem can be mapped to an equivalent channel model. Then, using random coding and typical set decoding [16] we upper bound the error probability, i.e., the probability of misclassifying the defective set.

#### Random matrix generation and the encoding process:

The binary measurement matrix is generated i.i.d. according to a bernoulli distribution. Associated with each item is a codeword that represents its assignment to tests. The defective set corresponds to a collection of  $K$  such codewords. Denote a collection of  $K$  codewords by  $X_{(K)}^T$  which is  $T$  independent copies of a collection of vectors  $X_{(K)}$ . A collection of  $K$  codewords and the observed  $y^T$  have a joint distribution

$$p(x_{(K)}^T, y^T) = \prod_{t=1}^T p(x_{(K)}(t))p(y(t)|x_{(K)}(t))$$

Before we describe our decoder, we define joint typicality in the context of this group testing problem.

*Definition 3.1:* A collection of  $K$  codewords (matrix of size  $K \times T$ ) and tests' outcomes  $y^T$  are said to be jointly typical if they belong the typical set  $A$ , which is properly defined in Eq.45 for the noiseless and additive noise models, and Eq. 49 for the dilution model.

#### Decoder:

Decoding is achieved using typical set decoding [16]. The decoder goes through all  $\binom{N}{K}$  possible sets of size  $K$ , where  $K$  is the size of the defective set, until it finds a set such that  $(x_{(K)}^T, y^T)$  are jointly typical (Definition 3.1). An error occurs if the generating set of codewords (corresponding to the true defective set) and the output  $y$  are not jointly typical, or if any set, other than the positive set, and  $y$  are jointly typical.

#### Analysis of the probability of error:

Given the random codebook generation let  $P_e$  denote the average probability of error, averaged over all codebooks and over all sets of size  $K$ , i.e.,:

$$\begin{aligned} P_e &= \sum_{\mathcal{C}} P(\mathcal{C}) \lambda(\mathcal{C}) \\ &= \sum_{\mathcal{C}} P(\mathcal{C}) \frac{1}{\binom{N}{K}} \sum_v \lambda_v \\ &= \frac{1}{\binom{N}{K}} \sum_v \sum_{\mathcal{C}} P(\mathcal{C}) \lambda_v(\mathcal{C}) \end{aligned} \tag{7}$$

By symmetry of the codebook construction,  $\sum_{\mathcal{C}} P(\mathcal{C})\lambda_v(\mathcal{C})$  does not depend on the set  $v$ . Thus,

$$P_e = \sum_{\mathcal{C}} P(\mathcal{C})\lambda_v(\mathcal{C}) = P_{e|v} \quad (8)$$

In other words, the average error probability does not depend on the input  $v$  due to averaging over randomly generated codebooks.

Now define the error event  $E_i$  as follows:

$E_i$ : The event that a set which differs in exactly  $i$  items from the true set is jointly typical with  $y$ . The probability of such an event is denoted  $P(E_i)$ .

Using the union bound, the average probability of error can now be bounded as:

$$P_e = P_{e|v} \leq P(E_0^c) + \sum_{i=1}^K P(E_i) \quad (9)$$

The first term corresponds to the event that the output  $y$  and the input to the channel (i.e. the codewords of the true defective set) are not jointly typical. In the appendix, given proper definition of typicality, we show that the probability of this event goes to zero fast for sufficiently large  $T$  and is dominated by the probabilities of the  $E_i$  events for all the models considered in this paper. See the appendix for the proof of an upper bound on the probability of the atypical event for the noise free, additive, and dilution models which we state in lemmas 4.1, 6.1, and 6.3, respectively.

We are interested in determining the required number of tests  $T$  to achieve an arbitrarily small error probability for large  $N$  and  $K$  with  $K \ll N$  and we choose  $p = \frac{1}{K}$ .

Before we bound the sum in Eq.(9) we prove the following lemma:

*Lemma 3.1:* If  $X_{(K)}$  and  $Y$  are jointly distributed as  $p(x_{(K)}, y)$  and  $\tilde{X}_{(K)}^T = (\tilde{X}_{(i)}^T, X_{(K-i)}^T)$  is a set of  $K$  codewords formed by replacing  $i$  of the codewords of the true defective set with  $i$  independent codewords (hence also independent of  $Y^T$ ) and keeping the remaining  $K - i$  codewords (with the same marginals), then the probability that  $(\tilde{X}_{(K)}^T, Y^T)$  is jointly typical can be bounded as:

$$\Pr[(\tilde{X}_{(i)}^T, X_{(K-i)}^T, Y^T) \in A_{\epsilon}^T] \leq 2^{-TI(X_{(i)}; X_{(K-i)}, Y)} \quad (10)$$

where  $A_{\epsilon}^T$  is the set of jointly typical sequences  $(X_{(K)}^T, Y^T)$  [16].

Recall that  $X_{(i)}$  is a collection of code symbols corresponding to a given set of  $i$  defective items. By symmetry of the codebook construction, the mutual information expression above depends only on the number of items  $i$  and not on which  $i$  items.

*Proof:*

$$\begin{aligned}
\Pr[(\tilde{X}_{(i)}^T, X_{(K-i)}^T, Y^T) \in A_\epsilon^T] &= \sum_{(x_{(K)}^T, y^T) \in A_\epsilon^T} P(x_{(i)}^T) P(x_{(K-i)}^T, y^T) \\
&\leq 2^{TH(X_{(K)}, Y)} 2^{-TH(X_{(i)})} \cdot 2^{-TH(X_{(K-i)}, Y)} \\
&= 2^{-TI(X_{(i)}; X_{(K-i)}, Y)}
\end{aligned} \tag{11}$$

■

Intuitively this means that the probability of obtaining jointly typical sequences, by replacing  $i$  codewords of the true set with  $i$  independent codewords, scales exponentially with the negative of the mutual information between  $i$  codewords of the set, and the remaining  $K - i$  codewords and the output. Before stating our main theorem, it is worthwhile mentioning that: In the classical channel coding problem [16], the error probability analysis for typical set decoding separates nicely due to the independence of the channel output and every codeword other than the truly transmitted one. However, in the group testing problem the main difficulty arises from the fact that a collection of  $K$  codewords and the true defective set could be overlapping. Hence, independence of the output and the collection of codewords does not hold anymore. To analyze the error probability, our approach was to isolate the overlap ensuring the independence of the remaining codewords from the true defective set and the output.

*Theorem 3.2:* Given  $N$  items with  $K$  unknown defectives, the following number of tests  $T$  is achievable for the noise-free and the noisy models of Eq.(2),(3) and (4),

$$T = \max_i \left\{ \frac{\log \binom{N-K}{i} \binom{K}{i}}{I(X_{(i)}; X_{(K-i)}, Y)} \right\} \tag{12}$$

where  $X_{(w)}$  denotes a collection of code symbols of size  $w$  (a vector) and  $Y$  denotes the observation.

Intuitively, the numerator represents the number of bits required to describe how many sets we can form with  $i$  misclassified items ( $i$  out of  $K$ ) and the denominator represents the amount of information if  $K - i$  of the defective items are known.

*Proof:* We can bound the probability of the event  $E_i$  by summing over all sets that differ from the true defective set in  $i$  codewords. We can form as many as  $\binom{K}{i} \binom{N-K}{i}$  such sets. Thus using Lemma 3.1,

$$P(E_i) \leq \binom{K}{i} \binom{N-K}{i} 2^{-TI(X_{(i)}; X_{(K-i)}, Y)} \tag{13}$$

We can now bound the total error probability for  $T$  sufficiently large as :

$$\begin{aligned}
P_e &\leq P(E_0^c) + \sum_i P(E_i) \\
&\leq P(E_0^c) + \sum_{i=1}^K \binom{K}{i} \binom{N-K}{i} 2^{-TI(X_{(i)}; X_{(K-i)}, Y)} \\
&= P(E_0^c) + \sum_{i=1}^K \binom{K}{i} \binom{N-K}{i} 2^{-T(H(Y|X_{(K-i)}) - H(Y|X_{(K)}))}
\end{aligned} \tag{14}$$

As we pointed out earlier and detailed the proof in the appendix (Lemmas 4.1, 6.1, 6.3), the probability of the atypical event  $P(E_0^c)$  is always dominated by the sum of  $P(E_i)$ . Hence  $T$  in Thm. 3.2 is achievable. ■

**Main Intuition:** The main idea behind the described approach is illustrated in Fig. 3. The group testing problem has been mapped to an equivalent multiple channel model. Each channel accounts for the case, where  $K - i$  of the defective items are recovered and  $i$  items are still to be recognized. The capacity of each channel contributes partially to the total error probability in decoding the true defective set.  $K - i$  represents the overlap between the true defective set and a false candidate set. By isolating the overlap, independence of the remaining  $i$  codewords from the codewords of the true set, helps bounding the associated error probability.

The result is a simple mutual information expression that can be used to determine the tradeoffs between  $K$ ,  $N$ ,  $T$ , and noise for various models as we show in the following sections. Table I summarizes the scaling results for the considered models for the average error and worst case error criteria. Note that the number of tests increases by  $1/(1 - u)^2$  factor for the dilution model and only by  $1/(1 - q)$  for the additive noise model which matches the aforementioned intuition.

#### IV. NOISE FREE CASE-DETERMINISTIC OUTPUT

In this section, we consider the noise-free (deterministic) case: the test outcome  $Y$  is 1 if and only if a defective item is pooled in that test. Hence  $Y$  is given by Eq. 1. We consider two scenarios: average and worst-case error. As mentioned earlier, the former requires the average error to be small and the latter considers the worst-case input since bounding the average error probability does not guarantee error-free performance for all possible defective sets.

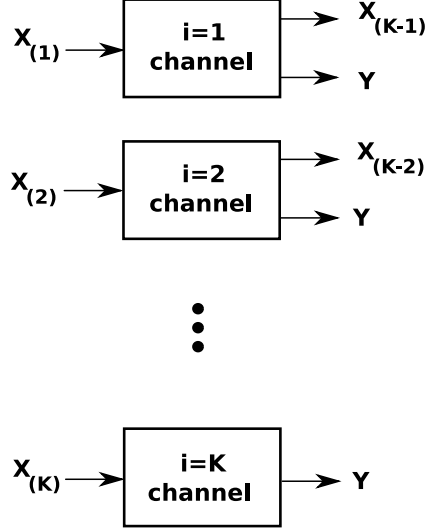


Fig. 3. Equivalent channel model. Each channel accounts for the case, where  $K - i$  of the defective items are recovered and  $i$  items are still to be recognized. The capacity of each channel contributes partially to the total error probability in decoding the true defective set.  $K - i$  represents the overlap between the true defective set and a false candidate set. By isolating the overlap, independence of the remaining  $i$  codewords from the codewords of the true set, helps bounding the associated error probability.

#### A. Average Error

Before we state the theorem, the following lemma provides an upper bound on the probability of the atypical event for the noise-free case. The proof is provided in the appendix.

*Lemma 4.1:* For the noiseless model of Eq.(1),  $\forall \delta > 0$ , there exists a constant  $c(\delta) > 0$  such that, for all  $T \geq c \cdot K \log K$ , the probability of the atypical event  $P(E_0^c)$  (Eq.9) (i.e., the probability that the outcome of the tests and the codewords of the true defective set are not jointly typical) is less than or equal to  $\delta$ .

Now we state the following theorem:

*Theorem 4.2:* For  $N$  items and  $K$  defectives, the number of tests  $T = O(K \log N)$  is sufficient to satisfy an average error criterion, i.e., achieve an arbitrarily small average error probability.

*Proof:* In the noise free case,  $H(Y|X_{(K)}) = 0$  since the outcome of a test is deterministic given the knowledge of all the items in the defective set. Recalling that  $X_{(i)}$  represents the code symbols of a set of  $i$  defective items, the mutual information expression  $I(X_{(i)}; X_{(K-i)}, Y)$  can be written as:

$$\begin{aligned}
I(X_{(i)}; X_{(K-i)}, Y) &= H(Y|X_{(K-i)}) - H(Y|X_{(K)}) \\
&= (1-p)^{K-i} H((1-p)^i) \\
&\geq \left(1 - \frac{1}{K}\right)^{K-i} \left(1 - \frac{1}{K}\right)^i \log \frac{1}{(1 - \frac{1}{K})^i}
\end{aligned} \tag{15}$$

Thus, for large  $K$ ,

$$I(X_{(i)}; X_{(K-i)}, Y) \geq e^{-1} \frac{i}{K \ln 2}$$

Using Thm.3.2 and the fact that  $\binom{N}{K} \leq 2^{NH(\frac{K}{N})}$  we can bound  $T$  as follows:

$$\begin{aligned}
T &\leq \frac{(N-K) \left[ \frac{i}{N-K} \log \frac{N-K}{i} + (1 - \frac{i}{N-K}) \log \frac{1}{1 - \frac{i}{N-K}} \right] + K \left[ \frac{i}{K} \log \frac{K}{i} + (1 - \frac{i}{K}) \log \frac{1}{1 - \frac{i}{K}} \right]}{e^{-1} \frac{i}{K}} \\
&\leq \frac{i \log N + \frac{1}{\ln 2} (N-K) (1 - \frac{i}{N-K}) (\sum_{j=1}^{\infty} \frac{1}{j} (\frac{i}{N-K})^j) + i \log K + \frac{1}{\ln 2} K (1 - \frac{i}{K}) (\sum_{j=1}^{\infty} \frac{1}{j} (\frac{i}{K})^j)}{e^{-1} \frac{i}{K}} \\
&= \frac{i \log N + \frac{i}{\ln 2} - \frac{1}{\ln 2} \sum_{j=2}^{\infty} \frac{1}{j(j-1)} \frac{i^j}{(N-K)^{j-1}} + i \log K + \frac{i}{\ln 2} - \frac{1}{\ln 2} \sum_{j=2}^{\infty} \frac{1}{j(j-1)} \frac{i^j}{K^{j-1}}}{e^{-1} \frac{i}{K}} \\
&\leq e \ln 2 (K \log(NK) + 2/\ln 2)
\end{aligned} \tag{16}$$

The second inequality uses the Taylor expansion of  $\log(1/(1-x))$ .

Hence,  $T = O(K \log(N))$ .

**Intuitive explanation:** The average error probability is bounded by the sum of probabilities of events  $E_i$ , where

$$\begin{aligned}
P(E_i) &= \binom{K}{i} \binom{N-K}{i} 2^{-T(H(Y|X_{(K-i)}))} \\
&= \binom{N-K}{i} \binom{K}{i} 2^{-T(1-p)^{K-i} H((1-p)^i)} \\
&\leq 2^{NH(\frac{i}{N-K})} \cdot 2^{KH(\frac{i}{K})} 2^{-T(1-p)^{K-i} H((1-p)^i)}
\end{aligned} \tag{17}$$

To understand the behavior of the average error probability, we study the behavior of the terms above for every value of  $i$ . We consider the following,  $i = o(K)$  and  $i = O(K)$ . We also choose  $p = \frac{1}{K}$ .

- First,  $i = o(K)$ . In this case:

$$\begin{aligned}
P(E_i) &\doteq 2^{i \log \frac{N}{i}} \cdot 2^{i \log \frac{K}{i}} \cdot 2^{-T(1 - \frac{1}{K})^{K-o(K)} H((1 - \frac{1}{K})^{o(K)})} \\
&\doteq 2^{i \log N} \cdot 2^{-TH(\frac{i}{K})} \doteq 2^{i \log N} \cdot 2^{-T\beta \frac{i}{K} \log \frac{K}{i}}
\end{aligned} \tag{18}$$

where  $\beta$  is a constant. Replacing for  $i = o(K)$  we could see that this term goes to zero fast for  $T = O(\frac{K \log N}{\log K})$

- Second considering the case where  $i = O(K)$ .

$$P(E_i) = \doteq 2^{NH(\frac{K}{N})} \cdot 2^{-T\beta} \quad (19)$$

$T = O(K \log N)$  is hence a sufficient condition for these terms.

■

### B. Maximum Probability of error

The previous analysis considered the average error case. Maintaining the average error probability below  $\epsilon$  is not enough if we are interested in the worst-case input, i.e. maximum error case. For exact reconstruction, the worst-case error is required to be zero.

*Theorem 4.3:* For  $N$  items and  $K$  defectives,  $T = O(K^2 \log N)$  is achievable for exact reconstruction (worst-case error criteria).

*Proof:* Since the average error probability is below  $\epsilon$ , then:

$$P_e = \sum_{\mathcal{C}} \Pr[\mathcal{C}] \lambda(\mathcal{C}) < \epsilon \rightarrow \exists \mathcal{C} : \lambda(\mathcal{C}) < \epsilon \quad (20)$$

to say that, since the average probability of error (over codebooks and inputs) is below  $\epsilon$ , then there exist a codebook  $\mathcal{C}$  such that  $\lambda(\mathcal{C}) = \frac{1}{\binom{N}{K}} \sum_v \lambda_v(\mathcal{C}) < \epsilon$ . Choosing  $\epsilon = \frac{1}{\binom{N}{K}}$  guarantees that the worst case error is also 0 since for the noiseless case  $\lambda_v \in \{0, 1\}$ . Driving the error terms to zero requires:

$$2^{\log N - \frac{T \log K}{K}} < \frac{1}{\binom{N}{K}} \quad (21)$$

Hence  $T = O(K^2 \log N)$ .

■

## V. ACHIEVABILITY WITH DISTORTION

In this section we relax our goal. We are satisfied with recovering a large fraction of the defective items. In other words, we allow an approximate reconstruction [15] in the sense that if  $K$  candidates are declared, up to  $\alpha K$  misses are allowed ( $\alpha$  small).

*Theorem 5.1:* If  $N$  is the total number of items and  $K$  the size of the defective set, approximate reconstruction of the defective set, i.e. with up to  $\alpha K$  misses, is achievable with  $T = O(\frac{e^{(1-\alpha)} K \log N}{H(e^{-\alpha})})$ .

*Proof:* In this case, to drive the error terms to zero:

$$\begin{aligned} \binom{N-K}{\alpha K} \binom{K}{\alpha K} 2^{-T(1-\frac{1}{K})^{K(1-\alpha)} H((1-\frac{1}{K})^{\alpha K})} &< \frac{1}{\binom{N}{K}} \\ 2^{\alpha K \log \frac{N}{K}} \cdot 2^{-Te^{-(1-\alpha)} H(e^{-\alpha})} &< \frac{1}{\binom{N}{K}} \end{aligned} \quad (22)$$

Hence, for large  $K$  and  $p = \frac{1}{K}$ ,  $T = \frac{e^{(1-\alpha)} K \log N}{H(e^{-\alpha})}$  is achievable. ■

*Remark 5.1:* The fact that the achievable bound above scales only with  $K$  suggests a multistage approach where  $K$  items are declared followed by  $\alpha K$  items etc... For the second stage, with  $\alpha^2 K$  distortion, we require  $T = \frac{e^{(1-\alpha)} \alpha K \log N}{H(e^{-\alpha})}$  tests. Hence for the multistage, the total number of tests required is such that:

$$T \leq \frac{e^{(1-\alpha)} K \log N}{H(e^{-\alpha})(1-\alpha)} \quad (23)$$

## VI. NOISY GROUP TESTING

The derived upper bound in Thm. 3.2 is fairly general as it maps the group testing problem to an equivalent channel model. This does not restrict the model to the noise-free scenario and hence could also be used to account for different noisy versions of the problem. The question reduces to how easy it is to compute the mutual information expression. In this section, we derive sufficient conditions on the number of tests for two types of noisy channels. It is to be noted that the result could also be applied to other noise models.

### A. Additive Observation Noise

First, we consider the additive output model of Eq.3 where:

$$Y_t = \left( \bigvee_{i=1}^N X_{ti} R_i \right) \vee W_t$$

where  $W$  is Bernoulli( $q$ ). This model captures the possibility of probabilistic false alarms. This accounts for errors in blood tests or background wireless losses [5] etc.

Again, we first upper bound the probability of the atypical event. See the appendix for a proof of the following lemma.

*Lemma 6.1:* For the additive model of Eq.(3),  $\forall \delta > 0$ ,  $\exists c > 0$  such that,  $\forall T \geq c \cdot \max\{K \log K, \frac{K}{1-q}\}$ , the probability of the atypical event  $P(E_0^c)$  in Eq.(9) (i.e., the probability that the outcome of the tests and the codewords of the true defective set are not jointly typical) is  $\leq \delta$ .



Now we state the following theorem:

*Theorem 6.2:* For the additive noise model of Eq.3,  $N$  items,  $K$  defectives,  $T = O(\frac{K \log N}{1-q})$  is achievable, where  $q$  is the parameter of the bernoulli distribution of the binary noise.

**Note:** As  $q$  increases, the number of tests required to identify the defective set increases since the outcome of a pooling test becomes less reliable due to false alarms.

*Proof:* Note that for the considered additive model, knowing  $K - i$  defective items might not completely determine the test outcome. The test outcome  $Y$  remains uncertain if their corresponding code symbols  $X_{(K-i)}$  are all 0. The test is negative only if the code symbols  $X_{(i)}$  for the remaining  $i$  items, as well as the realization of the Bernoulli noise for that test, are all 0. Thus,

$$\begin{aligned} I(X_{(i)}; X_{(K-i)}, Y) &= H(Y|X_{(K-i)}) - H(Y|X_{(K)}) \\ &= (1-p)^{K-i} H((1-p)^i(1-q)) - (1-p)^K H(q) \\ &= (1-p)^{K-i} (H((1-p)^i(1-q)) - (1-p)^i H(q)) \end{aligned} \quad (24)$$

The first entropy term can be written as:

$$\begin{aligned} H((1-p)^i(1-q)) &= (1-p)^i(1-q) \log \frac{1}{(1-p)^i(1-q)} + (1 - (1-p)^i(1-q)) \log \frac{1}{1 - (1-p)^i(1-q)} \\ &= i \left(1 - \frac{1}{K}\right)^i (1-q) \log \left(\frac{1}{1-p}\right) + \left(1 - \frac{1}{K}\right)^i (1-q) \log \left(\frac{1}{1-q}\right) \\ &\quad + \sum_{j=1}^{\infty} \frac{1}{j \ln 2} \left(1 - \frac{1}{K}\right)^{ji} (1-q)^j \\ &\quad - \frac{1}{\ln 2} \left(1 - \frac{1}{K}\right)^i (1-q) \left[ \sum_{j=1}^{\infty} \frac{1}{j} \left(1 - \frac{1}{K}\right)^{ji} (1-q)^j \right] \end{aligned} \quad (25)$$

The second entropy term in Eq.(24) can be written as:

$$(1-p)^i H(q) = \left(1 - \frac{1}{K}\right)^i q \log \frac{1}{q} + \left(1 - \frac{1}{K}\right)^i (1-q) \log \frac{1}{1-q} \quad (26)$$

Subtracting Equations (25) and (26), and multiplying by  $(1-p)^{K-i}$  the mutual information expression simplifies to:

$$\begin{aligned}
I(X_{(i)}; X_{(K-i)}, Y) &= \left(1 - \frac{1}{K}\right)^{K-i} \left[ \frac{i}{\ln 2} \left(1 - \frac{1}{K}\right)^i (1-q) \left( \sum_{j=0}^{\infty} \frac{1}{jK^j} \right) + \frac{1}{\ln 2} \left(1 - \frac{1}{K}\right)^i (1-q) \right. \\
&\quad \left. - \frac{1}{\ln 2} \left( \sum_{j=2}^{\infty} \frac{1}{j(j-1)} \left(1 - \frac{1}{K}\right)^{ji} (1-q)^j \right) - \left(1 - \frac{1}{K}\right)^i q \log \frac{1}{q} \right] \\
&\geq \frac{i}{\ln 2} \left(1 - \frac{1}{K}\right)^K (1-q) \left( \sum_{j=0}^{\infty} \frac{1}{jK^j} \right) + \frac{1}{\ln 2} \left(1 - \frac{1}{K}\right)^K (1-q) \\
&\quad - \frac{1}{\ln 2} \left(1 - \frac{1}{K}\right)^K \left[ \sum_{j=0}^{\infty} \frac{(1-q)^j}{j(j-1)} \right] - \left(1 - \frac{1}{K}\right)^K q \log \frac{1}{q}
\end{aligned} \tag{27}$$

Now note that

$$\sum_{j=0}^{\infty} \frac{(1-q)^j}{j(j-1)} = 1 - q - q \ln 2 \log \frac{1}{q}$$

since:

$$\begin{aligned}
\frac{(1-q)^2}{2} + \frac{(1-q)^3}{6} + \dots &= (1-q)^2 \left(1 - \frac{1}{2}\right) + (1-q)^3 \left(\frac{1}{2} - \frac{1}{3}\right) + \dots \\
&= (1-q) \left(1 - q + \frac{(1-q)^2}{2} + \dots\right) - \left( \frac{(1-q)^2}{2} + \frac{(1-q)^3}{3} + \dots \right) \\
&= (1-q) \log \frac{1}{q} - \left( \log \frac{1}{q} - (1-q) \right) \\
&= 1 - q - q \ln 2 \log \frac{1}{q}
\end{aligned} \tag{28}$$

Replacing in Eq.(27) we get:

$$I(X_{(i)}; X_{(K-i)}, Y) \geq \frac{i}{K \ln 2} \left(1 - \frac{1}{K}\right)^K (1-q) \tag{29}$$

Hence,  $T \leq \frac{(e \ln 2) K \log N}{1-q}$ , i.e.,  $T = O\left(\frac{K \log N}{1-q}\right)$  is achievable. ■

*Remark 6.1:* Following the same argument in Section IV-B, it is not hard to see that the same scaling holds for the worst-case error criteria but replacing  $K$  with  $K^2$  as shown in table I. In this noisy setup, this would mean that the worst-case error probability goes to zero in contrast to the noise-free case where this scaling ensures exact recovery.

### B. Dilution

The second noisy model we consider is the “dilution” model. Positive items taking part in a given test might probabilistically behave as absent (diluted). If all positive items in a given test appear as absent

that could potentially lead to erroneously zero outcomes. This model is motivated by blood dilution due to pooling with other negative tests, or imperfectly diluted blood samples, or adversarial camouflage-in the form of probabilistic transmission-in communication systems [5]. This is captured by the Z-channel model of Fig. 2. For this case, we show that  $T = O(\frac{K \log N}{(1-u)^2})$  is achievable. Intuitively, a larger flip probability  $u$  implies that more items will get diluted. As the tests become less reliable, a larger number of tests is needed to identify the defective set.

First, it is shown that the probability of the atypical event is dominated by the probability of the other error events (see Eq.9), which follows from the following lemma.

*Lemma 6.3:* For the dilution model of Eq.(4),  $\forall \delta > 0, \exists c > 0$  such that,  $\forall T \geq c \cdot K \log^2 K$ , the probability of the atypical event (Eq.9)  $P(E_0^c) \leq \delta$ .

*Proof:* See appendix ■

*Theorem 6.4:* Considering  $N$  items,  $K$  defectives and the dilution model represented by the Z-channel in Fig.(2) and Eq.(4),  $T = \frac{K \log N}{(1-u)^2}$  is achievable, where  $u$  is the transition probability of the z-channel (i.e. the probability that 1 is flipped into 0).

*Proof:* For notational convenience we let  $u = 1 - s$ . Recalling that  $X_{(i)}$  refers to the code symbols of a set of  $i$  defective items, the mutual information can be written as:

$$\begin{aligned} I(X_{(i)}; X_{(K-i)}, Y) &= H(Y|X_{(K-i)}) - H(Y|X_{(K)}) \\ &= \sum_{j=0}^{K-i} \binom{K-i}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-i-j} H\left(\sum_{\ell=0}^i \binom{i}{\ell} (1-s)^{j+\ell} \left(\frac{1}{K}\right)^\ell \left(1 - \frac{1}{K}\right)^{i-\ell}\right) \\ &\quad - \sum_{j=1}^K \binom{K}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-j} H((1-s)^j) \end{aligned} \quad (30)$$

The first sum, i.e.  $H(Y|X_{(K-i)})$ , can be written as:

$$\begin{aligned} H(Y|X_{(K-i)}) &= \sum_{j=0}^{K-i} \binom{K-i}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-i-j} H\left((1-s)^j \left(1 - \frac{s}{K}\right)^i\right) \\ &= \sum_{j=0}^{K-i} \binom{K-i}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-i-j} \left[ (1-s)^j \left(1 - \frac{s}{K}\right)^i \log \frac{1}{(1-s)^j \left(1 - \frac{s}{K}\right)^i} \right] \\ &\quad + \sum_{j=0}^{K-i} \binom{K-i}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-i-j} \left[ \left(1 - (1-s)^j \left(1 - \frac{s}{K}\right)^i\right) \log \frac{1}{1 - (1-s)^j \left(1 - \frac{s}{K}\right)^i} \right] \\ &= \eta_1 + \eta_2 \end{aligned} \quad (31)$$

Notice that:

$$\sum_{j=0}^{K-i} \binom{K-i}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-i-j} (1-s)^j = \left(1 - \frac{s}{K}\right)^{K-i} \quad (32)$$

$$\sum_{j=0}^{K-i} j \cdot \binom{K-i}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-i-j} (1-s)^j = \frac{1}{K} (K-i)(1-s) \left(1 - \frac{s}{K}\right)^{K-i-1} \quad (33)$$

Using Eq. (32) and (33), the first term  $\eta_1$  in Eq. (31) simplifies to:

$$\eta_1 = i \left(1 - \frac{s}{K}\right)^K \log\left(\frac{1}{1 - \frac{s}{K}}\right) + \frac{1}{K} (K-i) \left(1 - \frac{s}{K}\right)^{K-1} (1-s) \log \frac{1}{1-s} \quad (34)$$

Equivalently, the second sum, i.e.  $H(Y/X_{(K)})$  simplifies to:

$$\begin{aligned} H(Y/X_{(K)}) &= \sum_{j=0}^K \binom{K}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-j} (1-s)^j \log \frac{1}{(1-s)^j} \\ &\quad + \sum_{j=0}^K \binom{K}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-j} (1 - (1-s)^j) \log \frac{1}{1 - (1-s)^j} \\ &= \theta_1 + \theta_2 \end{aligned} \quad (35)$$

From Eq.(33),  $\theta_1$  simplifies to:

$$\theta_1 = \frac{1}{K} K (1-s) \left(1 - \frac{s}{K}\right)^{K-1} \log \frac{1}{1-s} \quad (36)$$

Combining  $\eta_1$  and  $\theta_1$  we get:

$$\begin{aligned} \eta_1 - \theta_1 &= i \left(1 - \frac{s}{K}\right)^{K-1} \left[ \left(1 - \frac{s}{K}\right) \log \frac{1}{1 - \frac{s}{K}} - \frac{1}{K} (1-s) \log \frac{1}{1-s} \right] \\ &= \frac{i}{\ln 2} \left(1 - \frac{s}{K}\right)^{K-1} \left[ \frac{s}{K} + O\left(\frac{1}{K^2}\right) - \frac{1}{K} (1-s) \left(s + \frac{s^2}{2} + \dots\right) \right] \\ &= \frac{i}{\ln 2} \left(1 - \frac{s}{K}\right)^{K-1} \left[ \frac{s^2}{K} \left(1 - \frac{1}{2}\right) + \frac{s^3}{K} \left(\frac{1}{2} - \frac{1}{3}\right) + \dots \right] \\ &\geq \frac{is^2}{2K \ln 2} \left(1 - \frac{s}{K}\right)^{K-1} \end{aligned} \quad (37)$$

Now we look at the remaining terms, i.e.  $\eta_2$  and  $\theta_2$ . It is sufficient to show that the difference  $\eta_2 - \theta_2$  is  $\geq 0$ :

$$\eta_2 = \sum_{j=0}^{K-i} \binom{K-i}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-i-j} \left[ \left(1 - (1-s)^j \left(1 - \frac{s}{K}\right)^i\right) \log \frac{1}{1 - (1-s)^j \left(1 - \frac{s}{K}\right)^i} \right]$$

$$= \frac{1}{\ln 2} \sum_{j=0}^{K-i} \binom{K-i}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-i-j} \left[1 - (1-s)^j \left(1 - \frac{s}{K}\right)^i\right] \left[\sum_{\ell=1}^{\infty} \frac{1}{\ell} (1-s)^{\ell j} \left(1 - \frac{s}{K}\right)^{\ell i}\right] \quad (38)$$

Using Eq.(32)  $\eta_2$  simplifies to:

$$\begin{aligned} \eta_2 &= \frac{1}{\ln 2} \left[ \left(1 - \frac{s}{K}\right)^K - \sum_{j=0}^{K-i} \binom{K-i}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-i-j} \left(\sum_{\ell=2}^{\infty} \frac{1}{\ell(\ell-1)} (1-s)^{\ell j} \left(1 - \frac{s}{K}\right)^{\ell i}\right) \right] \\ &= \frac{1}{\ln 2} \left[ \left(1 - \frac{s}{K}\right)^K - \sum_{\ell=2}^{\infty} \frac{1}{\ell(\ell-1)} \left(1 - \frac{s}{K}\right)^{\ell i} \left(1 - \frac{1}{K} + \frac{1}{K}(1-s)^{\ell}\right)^{K-i} \right] \end{aligned} \quad (39)$$

Similarly,

$$\begin{aligned} \theta_2 &= \frac{1}{\ln 2} \left[ \left(1 - \frac{s}{K}\right)^K - \sum_{j=0}^K \binom{K}{j} \left(\frac{1}{K}\right)^j \left(1 - \frac{1}{K}\right)^{K-j} \left(\sum_{\ell=2}^{\infty} \frac{1}{\ell(\ell-1)} (1-s)^{\ell j}\right) \right] \\ &= \frac{1}{\ln 2} \left[ \left(1 - \frac{s}{K}\right)^K - \sum_{\ell=2}^{\infty} \frac{1}{\ell(\ell-1)} \left(1 - \frac{1}{K} + \frac{1}{K}(1-s)^{\ell}\right)^K \right] \end{aligned} \quad (40)$$

Comparing equations (39) and (40) it is now clear that for large  $K$ ,  $\eta_2 - \theta_2 \geq 0$ . This is easy to verify since:

$$\left(1 - \frac{1}{K} + \frac{1}{K}(1-s)^2\right)^K = \left(1 - \frac{1}{K} + \frac{1}{K}(1-s)^2\right)^{K-i} \left(1 - \frac{1}{K} + \frac{1}{K}(1-s)^2\right)^i$$

Now,

$$\begin{aligned} \left(1 - \frac{1}{K} + \frac{1}{K}(1-s)^2\right) &= 1 - \frac{2s}{K} + \frac{s^2}{K} \\ &> \left(1 - \frac{s}{K}\right)^2 = 1 - \frac{2s}{K} + \frac{s^2}{K^2} \end{aligned}$$

■

Thus,  $T = O\left(\frac{K \log N}{s^2}\right)$  is achievable. Replacing for  $s = 1 - u$ , Thm. 6.4 follows.

*Remark 6.2:* Following the same argument in Section IV-B, it is not hard to see that the same scaling holds for the worst-case error criteria but replacing  $K$  with  $K^2$  as shown in table I.

## VII. LOWER BOUND: FANO'S INEQUALITY

In this section we also derive lower bounds on the required number of tests using Fano's inequality.

*Theorem 7.1:* For  $N$  items and  $K$  defectives, a lower bound on the total number of tests required to recover the defective set is given by:

$$T \geq \frac{\log \binom{N}{K}}{I(X_{(K)}; Y)} \quad (41)$$

*Proof:* The average error probability can be lower bounded using Fano's inequality:

$$\begin{aligned} P_e &\geq \frac{H(V|Y^T) - 1}{\log \binom{N}{K}} \\ &= 1 - \frac{I(V; Y^T) - 1}{\log \binom{N}{K}} \\ &\stackrel{(a)}{\geq} 1 - \frac{I(X_{(K)}^T; Y^T) - 1}{\log \binom{N}{K}} \end{aligned} \quad (42)$$

where (a) follows from the data processing inequality. Hence bounding the error probability requires:

$$\begin{aligned} \log \binom{N}{K} &\leq I(X_{(K)}^T; Y^T) \\ &\stackrel{(a)}{=} \sum_{t=1}^T H(Y(t)|Y_1^{t-1}) - H(Y(t)|X_{(K)}(t)) \\ &\stackrel{(b)}{\leq} \sum_{t=1}^T H(Y(t)) - H(Y(t)|X_{(K)}(t)) \\ &= \sum_{t=1}^T I(X_{(K)}(t); Y(t)) \\ &\leq T \max_{p(X)} I(X_{(K)}; Y) \end{aligned} \quad (43)$$

(a) is due to the fact that tests are memoryless and (b) is true since conditioning reduces entropy. Hence it is necessary that  $T \geq \frac{\log \binom{N}{K}}{\max_{p(X)} I(X_{(K)}; Y)}$ . ■

*Remark 7.1:* According to this lower bound, it is not hard to see that for the noise free case  $T \geq K \log N$  is a necessary condition on the total number of tests.

*Remark 7.2:* It is straightforward to verify that the lower bound depends on the minimum term in the upper bound summation since:

$$\begin{aligned} \sum_i \binom{N-K}{i} \binom{K}{i} 2^{-TI(X_{(i)}; X_{(K-i)}, Y)} &\stackrel{(1)}{\geq} \sum_i \binom{N-K}{i} \binom{K}{i} 2^{-TI(X_{(K)}; Y)} \\ &= \binom{N}{K} 2^{-TI(X_{(K)}; Y)} \end{aligned} \quad (44)$$

(1) is true since conditioning reduces entropy which ensures that,

$$I(X_i; X_{(K-i)}, Y) = H(Y/X_{(K-i)}) - H(Y/X_{(K)}) \leq H(Y) - H(Y/X_{(K)}) = I(X_{(K)}; Y)$$

It is clear that the number of tests required to let the R.H.S. of Eq. (44) go to zero matches the derived lower bound, i.e.  $T \geq \frac{\log \binom{N}{K}}{I(X_{(K)}; Y)}$ .

### VIII. CONCLUSIONS

In this paper, we adopted a new information theoretic framework to address group testing problems. This approach shifts the philosophy of random disjunct/separable matrix generation to an equivalent channel model and capacity computation. The result is a fairly general achievable bound that enables us to obtain the required tradeoffs between the number of tests, number of items and number of defective items for a wide range of group testing problems. Obtaining these tradeoffs reduces to a simple computation of mutual information expressions. We obtain the asymptotic scaling for i) noise-free setups with average and worst-case errors; ii) Approximate reconstruction and iii) Noisy versions of group testing, namely additive and dilution models.

### IX. ACKNOWLEDGMENTS

The authors would like to thank Anant Sahai at UC Berkeley and Nan Ma at Boston University for very useful discussions.

### APPENDIX

#### Proof of Lemma 4.1:

For both the noiseless and the additive noise models, define the typical set  $A$  as:

$$A = \left\{ (x^{K \times T}, y^T) : \begin{aligned} & \left| \frac{N(x_{(i)} = 0, y = 0)}{T} - p(x_{(i)} = 0, y = 0) \right| \leq \epsilon p(x_{(i)} = 0, y = 0), \\ & \left| \frac{N(x_{(i)} = 0, y = 1)}{T} - p(x_{(i)} = 0, y = 1) \right| \leq \epsilon p(x_{(i)} = 0, y = 1), \forall i \\ & \left| \frac{N(x_l)}{T} - p(x_l) \right| \leq \epsilon p(x_l), x_l \in \{0, 1\}, \forall l \end{aligned} \right\} \quad (45)$$

where  $N(\cdot)$  is the number of occurrences. Now we show that the probability of being atypical is bounded. The probability that the matrix of codewords corresponding to the true defective set and the tests outcomes are not jointly typical can be bounded as:

$$\begin{aligned}
P(A^c) &\stackrel{(a)}{\leq} \Pr \left[ \left| \frac{N(x_{(i)} = 0, y = 0)}{T} - p(x_{(i)} = 0, y = 0) \right| > \epsilon p(x_{(i)} = 0, y = 0) \text{ for some } i \right] \\
&+ \Pr \left[ \left| \frac{N(x_{(i)} = 0, y = 1)}{T} - p(x_{(i)} = 0, y = 1) \right| > \epsilon p(x_{(i)} = 0, y = 1) \text{ for some } i \right] \\
&+ \Pr \left[ \left| \frac{N(x_l)}{T} - p(x_l) \right| > \epsilon p(x_l) \text{ for some } l, x_l \in \{0, 1\} \right] + \\
&\stackrel{(b)}{\leq} 2^K \cdot 2^{-Tp(x_{(i)}=0, y=0)\epsilon^2} + \sum_i \binom{K}{i} 2^{-Tp(x_{(i)}=0, y=1)\epsilon^2} + K \cdot (2^{-Tp(0)\epsilon^2} + 2^{-Tp(1)\epsilon^2}) \quad (46)
\end{aligned}$$

where (a) is obtained by applying the union bound, (b) follows from the union and Chernoff bounds.

Now consider the noise free model. In this case,

$$\begin{aligned}
p(x_{(i)} = 0, y = 0) &= \left(1 - \frac{1}{K}\right)^K \\
p(x_{(i)} = 0, y = 1) &= \left(1 - \frac{1}{K}\right)^i \left(1 - \left(1 - \frac{1}{K}\right)^{K-i}\right)
\end{aligned}$$

Also note that in the noise free case the index  $i$  in the summation runs from 1 to  $K - 1$  since the  $K$ -th term is always zero. This is true because the boolean sum of an all zero vector cannot be 1. Hence, there always exists a constant  $c_1 > 0$  such that the first term in Eq.46 is arbitrarily close to zero for all  $T \geq \frac{c_1 K}{\epsilon^2 (1 - \frac{1}{K})^K}$ , i.e., as  $K \rightarrow \infty$ ,  $\forall T \geq \text{const} * K$ . Similarly, each term in the second summation is arbitrarily small for all  $T \geq c_2 K \log K$  for some constant  $c_2$ . Also  $\exists c_3 > 0$  such that the last term is arbitrarily small  $\forall T \geq c_3 K \log K$  since  $p(0) = \frac{1}{K}$ .

### Proof of Lemma 6.1

Now consider the additive noise model of Eq.(3). In this case,

$$\begin{aligned}
p(x_{(i)} = 0, y = 0) &= (1 - q) \left(1 - \frac{1}{K}\right)^K \\
p(x_{(i)} = 0, y = 1) &= \left(1 - \frac{1}{K}\right)^i \left(1 - (1 - q) \left(1 - \frac{1}{K}\right)^{K-i}\right)
\end{aligned}$$

Thus, the first term in Eq. 46 for this additive model is arbitrarily small for all  $T \geq \frac{c_1 K}{1-q}$  with a proper choice of a constant  $c_1$ . For all  $T \geq c_2 K \log K$ , the terms in the summation of the second term are arbitrarily small since the probability  $p(x_{(i)} = 0, y = 1)$  in the exponent of each term is in fact greater than for the noise free case.

Now we verify that the set  $A$  as defined provides us with the desired entropic properties. For the noise free and additive noise model we observe that for any set  $X_{(i)}$  we can write the joint entropy as:



$$\begin{aligned}
H(X_{(i)}, Y) &= - \sum_{x_{(i)}} \sum_y p(x_{(i)}, y) \log p(x_{(i)}, y) \\
&\stackrel{(1)}{=} -p(x_{(i)} = 0, y = 0) \log p(x_{(i)} = 0, y = 0) - p(x_{(i)} = 0, y = 1) \log p(x_{(i)} = 0, y = 1)
\end{aligned} \tag{47}$$

where (1) follows from the fact that having any of the entries of the vector  $X_{(i)}$  equals to 1 would make  $y$  deterministically 1 for these 2 models. Now we write the empirical joint distribution,

$$\begin{aligned}
-\frac{1}{T} \sum_{t=1}^T \log p(x_{(i),t}, y_t) &= -\frac{N(x_{(i)} = 0, y = 0)}{T} \log p(x_{(i)} = 0, y = 0) \\
&\quad - \frac{N(x_{(i)} = 0, y = 1)}{T} \log p(x_{(i)} = 0, y = 1)
\end{aligned} \tag{48}$$

Comparing Eqs.47 and 48 it is clear that the set  $A$  ensures that the empirical joint distributions converge to the true joint entropies, in other words, for any  $(x^{K \times T}, y) \in A$ ,  $p(x^{i \times T}, y) \approx 2^{-TH(X_{(i)}, Y)}$  for any sub-matrix of  $i$  rows. Note that given the last conditions in the definition of the typical set  $A$ , it is straightforward that the empirical marginal distribution  $-\frac{1}{T} \log p(x^{i \times T})$  approximates the entropy  $H(X_{(i)}), \forall i$ .

### Proof of Lemma 6.3

The dilution case is more involved and requires a different definition for the typical set to ensure that the entropic properties are all met. We adopt the conventional definition of weak typicality [16] with an extra constraint that the number of ones per column of  $x^{K \times T}$  (the  $t$ -th column is denoted  $x_{(k)}^t$ ), does not exceed  $O(\log K)$ . More specifically the typical set  $A$  is defined as,

$$\begin{aligned}
A = \left\{ (x^{K \times T}, y^T) : \left| -\frac{1}{T} \log p(x_{(i)}, y) - H(X_{(i)}, Y) \right| \leq \epsilon H(X_{(i)}, Y) \quad \forall i, \right. \\
\left. |x_{(k)}^t|_1 = O(\log K), \quad \forall t \right\}
\end{aligned} \tag{49}$$

Now let

$$Z_{ijt}(y) = \left( \mathbb{I}_{\{|x_{(i)}^t|_1=j, y_t=y\}} - \binom{i}{j} p(j, y) \right) \log p(j, y), y \in \{0, 1\}$$

where,  $\mathbb{I}_{\{\cdot\}}$  is an indicator function and

$$p(j, 0) = \left( \frac{u}{K} \right)^j \left( 1 - \frac{1}{K} \right)^{i-j} \sum_{\ell=0}^{K-i} \binom{K-i}{\ell} \left( \frac{u}{K} \right)^\ell \left( 1 - \frac{1}{K} \right)^{K-i-\ell}$$

and,

$$p(j, 1) = \left(\frac{u}{K}\right)^j \left(1 - \frac{1}{K}\right)^{i-j} \left(1 - \sum_{\ell=0}^{K-i} \binom{K-i}{\ell} \left(\frac{u}{K}\right)^\ell \left(1 - \frac{1}{K}\right)^{K-i-\ell}\right)$$

Note that  $p(j, y)$  corresponds to an instance where a column has exactly  $j$  ones. The first set of conditions in the definition of the typical set is equivalent to:

$$\left| \frac{1}{T} \sum_{t=1}^T (Z_{it}(0) + Z_{it}(1)) \right| \leq \epsilon H(X_{(i)}, Y), \forall i$$

where  $Z_{it}(y) = \sum_{j=0}^i Z_{ijt}(y)$ ,  $y \in \{0, 1\}$ . The second condition in the definition of the typical set means that being typical requires the number of ones per column to be upper bounded by  $\log K$ .

For a fixed  $i$ , the random variables  $\zeta_{it} = Z_{it}(0) + Z_{it}(1)$  are independent with zero mean. Thus, if  $|\zeta_{it}| \leq M$ , then Bernstein's inequality implies that:

$$\Pr\left[\sum_{t=1}^T \zeta_{it} > \beta\right] \leq \exp\left\{-\frac{\beta^2/2}{\sum_t E[\zeta_{it}^2] + M\beta/3}\right\} \quad (50)$$

Since  $|Z_{it}(y)| \leq (-\log \min p(j, y))$ , and the total number of ones in the measurement matrix is bounded by  $\log K$ , then  $M = O(\log^2 K)$ . Furthermore, it is not hard to verify that the variance of  $\zeta_{it}$  is  $O((\frac{iu}{K})^2)$ .

Replacing in Eq.(50) we can see that:

$$\begin{aligned} \Pr\left[\sum_{t=1}^T \zeta_{it} > \epsilon H(X_{(i)}, Y)\right] &\leq \Pr\left[\sum_{t=1}^T \zeta_{it} > \epsilon \frac{iu}{K} \log \frac{K}{u}\right] \\ &\leq \exp\left(\frac{-\frac{T^2}{2} \left(\frac{iu}{K}\right)^2 \log^2 \left(\frac{K}{u}\right)}{T \left(\frac{iu}{K}\right)^2 + T \frac{iu}{K} \log^2 K}\right) \end{aligned} \quad (51)$$

The inequality in the opposite direction, i.e.,  $\Pr\left[\sum_{t=1}^T \zeta_{it} < -\epsilon H(X_{(i)}, Y)\right]$ , follows the exact same analysis. Now, since the probability of having more than  $\log K$  ones per column goes to zero exponentially fast, then using the union bound we can upper bound the probability of being atypical:

$$\Pr[A^c] \leq \sum_i \binom{K}{i} \exp\left\{-\frac{\frac{iu}{K} T \log^2 \left(\frac{K}{u}\right)}{\log K \log \left(\frac{K}{u}\right)}\right\} \quad (52)$$

which is again arbitrarily close to zero with  $T \geq cK \log^2 K$ , where  $c$  is a constant.

## REFERENCES

- [1] George Atia and Venkatesh Saligrama, “Noisy group testing: An information theoretic perspective,” in *Proceedings of the 47th Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, 2006.
- [2] Ding-Zhu Du and Frank K. Hwang, *Combinatorial Group Testing and Its Applications*, World Scientific Publishing Company, second edition, 2000.
- [3] R. Dorfman, “The detection of defective members of large populations,” *Ann. Math. Stat.*, 1943.
- [4] George Atia, Anant Sahai, and Venkatesh Saligrama, “Spectrum enforcement and liability assignment in cognitive radio systems,” in *Proceedings of the IEEE DySPAN*, Chicago, IL, Oct. 2008.
- [5] George Atia, Venkatesh Saligrama, and Anant Sahai, “Codes to unmask spectrum violators,” in *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2008.
- [6] V. Saligrama and M. Zhao, “Thresholded basis pursuit: Support recovery for sparse and approximately sparse signals,” Available: <http://arxiv.org/abs/0809.4883>, 2008.
- [7] W. H. Kautz and R.C. Singleton, “Nonrandom binary superimposed codes,” *IEEE Trans. Inf. Theory*, vol. 10, no. 4, pp. 363–377, Oct. 1964.
- [8] A. G. Dyachkov and V. V. Rykov, “A survey of superimposed code theory,” *Problems of Control and Information Theory*, vol. 12, no. 4, pp. 1–13, 1983.
- [9] P. Erdős, P. Frankl, and Z. Füredi, “Family of finite sets in which no set is covered by the union of  $n$  others,” *Isr. J. Math.*, vol. 51, pp. 79–89, 1985.
- [10] Hong-Bin Chen and Frank K. Hwang, “Exploring the missing link among  $d$ -separable,  $t$ -separable and  $d$ -disjunct matrices,” *Discrete Applied Mathematics*, vol. 155, no. 5, pp. 662 – 664, 2007.
- [11] A. G. Dyachkov, V. V. Rykov, and A. M. Rashad, “Bounds of the length of disjunct codes,” *Problems Control Information Theory*, vol. 11, pp. 7–13, 1982.
- [12] M. Ruszinkó, “Family of finite sets in which no set is covered by the union of  $n$  others,” *Isr. J. Math.*, vol. 51, pp. 79–89, 1985.
- [13] Andras Sebo, “On two random search problems,” *Journal of Statistical Planning and Inference*, vol. 11, pp. 23–31, 1985.
- [14] T. Berger and V.I. Levenshtein, “Asymptotic efficiency of two-stage disjunctive testing,” *Information Theory, IEEE Transactions on*, vol. 48, no. 7, pp. 1741–1749, Jul 2002.
- [15] A. C. Gilbert, M. A. Iwen, and M. J. Strauss, “Group testing and sparse signal recovery,” 2008.
- [16] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, New York: John Wiley and Sons, Inc., 1991.